for wireless personal area networks (PANs). USB Controller **240** also provides USB connectivity to other miscellaneous USB connected devices **242**, such as a mouse, removable nonvolatile storage device **245**, modems, network cards, ISDN connectors, fax, printers, USB hubs, and many other types of USB connected devices. While removable nonvolatile storage device **245** is shown as a USB-connected device, removable nonvolatile storage device **245** could be connected using a different interface, such as a Firewire interface, etcetera.

[0033] Wireless Local Area Network (LAN) device **275** connects to Southbridge **235** via the PCI or PCI Express bus **272**. LAN device **275** typically implements one of the IEEE 802.11 standards of over-the-air modulation techniques that all use the same protocol to wireless communicate between information handling system **200** and another computer system or device. Optical storage device **290** connects to Southbridge **235** using Serial ATA (SATA) bus **288**. Serial ATA adapters and devices communicate over a high-speed serial link. The Serial ATA bus also connects Southbridge **235** to other forms of storage devices, such as hard disk drives. Audio circuitry **260**, such as a sound card, connects to Southbridge **235** via bus **258**. Audio circuitry **260** also provides functionality such as audio line-in and optical digital audio in port **262**, optical digital output and headphone jack **264**, internal speakers **266**, and internal microphone **268**. Ethernet controller **270** connects to Southbridge **235** using a bus, such as the PCI or PCI Express bus. Ethernet controller **270** connects information handling system **200** to a computer network, such as a Local Area Network (LAN), the Internet, and other public and private computer networks.

[0034] While FIG. **2** shows one information handling system, an information handling system may take many forms, some of which are shown in FIG. **1**. For example, an information handling system may take the form of a desktop, server, portable, laptop, notebook, or other form factor computer or data processing system. In addition, an information handling system may take other form factors such as a personal digital assistant (PDA), a gaming device, ATM machine, a portable telephone device, a communication device or other devices that include a processor and memory.

[0035] FIGS. **3-9** depict an approach that can be executed on an information handling system, to provide mine threaded online discussions for answers to questions based using a knowledge management system, such as QA System **100** shown in FIG. **1**. This approach mines threaded discussions for solutions to posed questions, and then correlates the discussion content across many threads from disparate sites to determine the most likely answers to questions, as an additional source of evidence for use within a deep question answering system. Those skilled in the art with appreciate the applicability of this approach in a deep question answering system where a common question might be posed by different users in different online discussion using varying phrasing. For example a question such as "How do I change my oil in my 95 Pontiac" and "Anyone know the best way to get my oil filter off my 94 Pontiac?" Both questions can be answered using the techniques and approached described herein. Answers can be pre-computed and stored for later use or calculated on-demand for a user's question. These techniques and approaches take into account the vast amount of crowd-sourced data contained across any number of online discussions (e.g., forums) in existence. In addition, this approach leverages the evidence-strengthening aspect of finding the same answer in multiple online discussions. The approach applies novel language processing techniques to normalize the utterances into a form appropriate for use within a question-answering system. As described in further detail herein, threaded discussions have a variety of special properties that can be utilized in evidence scoring. For example, many online discussions include "ratings" of the answer posts provided by forum readers, site reputation, etc. In addition, other posts found in discussion threads, such as confirming that a posed answer is correct, can be identified and used as supporting evidence for a given candidate answer. In addition, sentiment analysis of responses to each answer post are utilized, where sentiment phrases indicate positivity or negativity toward the post. Threaded discussions may also contain dynamic up-to-the-minute information, compared to other document sources such as written web pages or technical manuals, which might not be frequently updated. Other properties found in online threaded discussions, such as the length/perplexity of an answer thread can also provide useful information to the question answering system. For example, if a particular question initiates complex or lengthy discussions or is answered with a wide variety of answers from other discussion participants, the question answering system might be able to conclude that the question does not have a definitive answer. Once identified, the candidate answers are scored using a variety of factors to select the most likely correct answer to a question. Some embodiments present a set of candidate answers where the set could be one answer or a number of candidate answers.

[0036] In contrast to other types of data ingested by a QA System, threaded discussions provide posts that can come from a number of different users. The new information is generally contained in a different post at a different date. The intersection of different authors, with each author likely having a different reliability rating, and time related factors. For example, newer posts could be more accurate about current conditions, however contemporaneous posts might be more accurate in terms of the information that the original poster to the threaded discussion requested in their question.

[0037] FIG. **3** is a component diagram depicting the various components in mining threaded online discussions. Online threaded discussion sites **300** can include any number of web sites with online threaded discussions, such as forums or the like, to which users pose questions and other users attempt to provide guidance and answers to the posed questions. In one embodiment, two phases are used to mine threaded online discussions. Discussion ingestion phase **310** is used to analyze threaded discussions from threaded discussion sites **300** and update a corpus (corpus **340**) that is utilized by the question answering system. The question answering phase is handled by question answering pipeline **350** that utilizes the corpus provided by the discussion ingestion phase in order to identify candidate answers as well as to select the most likely answer from the candidate answers.

[0038] Discussion ingestion **310** commences with harvest discussions process **315** that harvests the threaded discussions and stores the harvested discussions in harvested discussions data store **320**. Next, classify posts process **325** is used to classify the posts within the harvested discussions data store in order to build a set of classified posts found in the threaded discussions which are shown being stored in